



SP-SP

Working Paper
WP no 624
March, 2006

University of Navarra

**MEASURING INDUSTRY-SCIENCE LINKS THROUGH
INVENTOR-AUTHOR RELATIONS:
A PROFILING METHOD**

Bruno Cassiman

Patrick Glenisson

Bart Van Looy

IESE Business School – Universidad de Navarra

Avda. Pearson, 21 – 08034 Barcelona, España. Tel.: (+34) 93 253 42 00 Fax: (+34) 93 253 43 43

Camino del Cerro del Águila, 3 (Ctra. de Castilla, km 5,180) – 28023 Madrid, España. Tel.: (+34) 91 357 08 09 Fax: (+34) 91 357 29 13

Copyright © 2006 IESE Business School.

The Public-Private Center is a Research Center based at IESE Business School. Its mission is to develop research that analyses the relationships between the private and public sectors primarily the following areas: regulation and competition, innovation, regional economy and industrial politics and health economics.

Research results are disseminated through publications, conferences and colloquia. These activities are aimed to foster cooperation between the private sector and public administrations, as well as the exchange of ideas and initiatives.

The sponsors of the SP-SP Center are the following:

- Accenture
- Ajuntament de Barcelona
- Official Chamber of Commerce, Industry and Navigation of Barcelona
- BBVA
- Diputació de Barcelona
- Garrigues, Abogados y Asesores Tributarios
- Catalan Government (Generalitat de Catalunya)
- Sanofi-Aventis
- Telefonica
- T-Systems
- VidaCaixa

The content of this publication reflects the conclusions and findings of the individual authors, and not the opinions of the Center's sponsors.

MEASURING INDUSTRY-SCIENCE LINKS THROUGH INVENTOR-AUTHOR RELATIONS: A PROFILING METHOD

Bruno Cassiman*

Patrick Glenisson**

Bart Van Looy***

Abstract

In this pilot study we examine the performance of text-based profiling in recovering a set of validated inventor-author links. In a first step we match patents and publications based solely on their similarity in content. Next, we compare inventor and author names on the highest ranked matches for the occurrence of name matches. Finally, we compare these candidate matches with the names listed in a validated set of inventor-author names. Our text-based profiling method performs significantly better than random matching of patents and publications, suggesting that text-based profiling is a valuable complementary tool to the name searches used in previous studies.

* Professor, General Management, IESE and KU Leuven

** Professor, KU Leuven and Business & Decision Benelux

*** Professor, Managerial Economics, Strategy and Innovation, Faculty of Economics and Applied Economics, KU Leuven & Steunpunt O&O Statistieken

Keywords: innovation, industry-science links, text-based profiling

MEASURING INDUSTRY-SCIENCE LINKS THROUGH INVENTOR-AUTHOR RELATIONS: A PROFILING METHOD*

Introduction

In our current knowledge-driven society the dichotomous and unidirectional account of the interplay between science and technology has become inadequate. Indeed, just as universities and research institutions engage in entrepreneurial activities, we observe increased commitments with industry players to expand their science base (Gibbons et al., 1994; Etzkowitz and Leydesdorff, 1997; Meyer-Krahmer and Schmoch, 1998). The Triple Helix conferences illustrate the importance of this observation. Bibliometric or informetric analysis of patents and publications are attractive ways to model science-technology interaction. Several approaches to track science-technology interactions have been proposed (see Moed et al., 2004, and the contributions it contains). On the one hand, patent grants or applications are analysed for citations to the scientific literature or for scientific institutions as the patent assignee. Publications, on the other hand, are analysed for citations to patents, for authoring by industry, or for co-authoring relationships between university and industry. In this contribution we focus on a more recent development in directly linking patents and publications, while searching for inventor-author links (Rabeharisoa, 1992; Noyons et al., 1994; Van Looy et al., 2006; Schmoch, 2004; Meyer, 2005). We attempt to extend this approach by means of (content-based) lexical matching procedures.

Inventor-author links – instances of researchers who generate (basic) research insights and at the same time are actively developing technology – can help identify points of exchange between academic research and technology development. Schmoch (2004) shows that accounting for such links in the field of genetics increases the patents related to universities from 30% of patents with a university as assignee to 52% of patents with university employees as inventors. Meyer (2005), in a study of the field of nanotechnology and nanoscience, shows that inventors who also publish are relatively more productive and more cited than the average scientist working in the field.

* All authors are researchers at the Steunpunt O&O Statistieken of the KU Leuven and thank Arnold Verbeek for excellent research assistance. Bruno Cassiman acknowledges financial support from the European Commission Key Action “Improving the socio-economic knowledge base” through contract No. HPSE-CT-2002-00146, the Public-Private Sector Research Center at IESE Business School and the Spanish Government research grant (SEC 2003-08282).

In this paper we demonstrate how a pure lexical matching method can support the uncovering of links between “authors” of scientific publications and “inventors” of patents. To achieve this connection we link two different databases: the Web of Science (WoS by ISI-Thomson) publication database, which features academic and scientific publications, and the European patenting system (EPO) database covering all patents issued or granted. Our method involves mapping and coupling the activity profiles of authors and inventors by creating content-based matches (between authors and inventors) based on similarities between (text) profiles derived from patent documents and published articles. In what follows we describe and illustrate our lexical matching approach and demonstrate its validity by assessing its recognition performance in a well-defined case. In addition to showing proof-of-concept, we discuss the potential and limitations of our method and suggest how it can be extended for larger-scale applications.

Current Profiling Methods

Existing profile matching first extracts a reference set of author (inventor) names from university records (Meyer et al., 2003; Van Looy et al., 2006; Rapmund et al., 2004) or a preset definition of a science and technology field (Rabehariso, 1992, in fuel cells; Noyons et al., 1994, in laser medicine; Schmoch, 2004, in genetics; Meyer, 2005, in nano-science and technology). Next, an algorithm matches authors and inventors in the reference set based on last name, followed by a number of additional criteria such as initials or full first name, address information of the inventor and author, and scientific or technological field of publication and patenting.

This matching method, however, poses several important challenges. First, candidate matches generated by the profiling method might not constitute true matches and so may result in false positives. This is the case with “homonyms”, i.e. same name for author and inventor, but referring to a different physical person. The broader the reference set, the more likely the occurrence of such false positives. Additional controls, such as the address field, to discriminate between these cases are unlikely to be very powerful, in particular when analyzing the name list of a particular university, where all researchers are located close to each other.

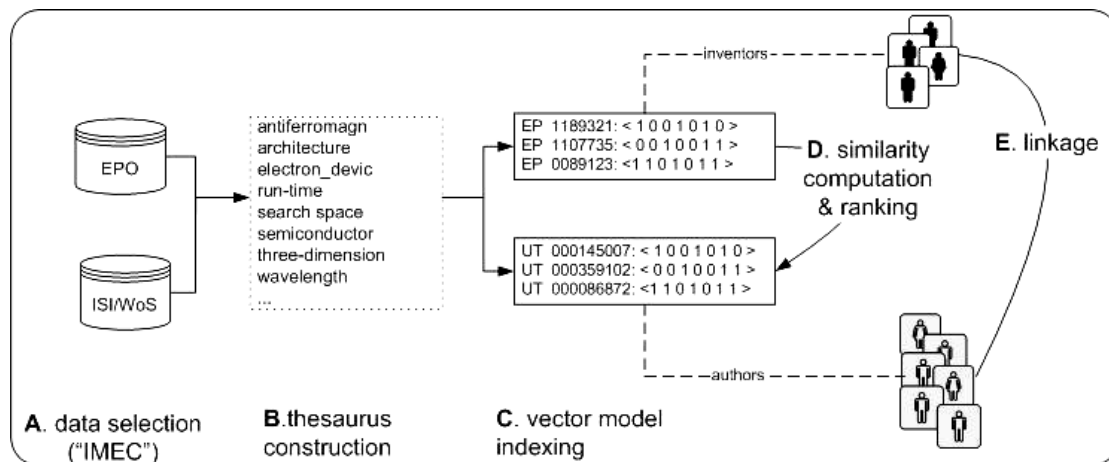
A second and potentially more important problem is the fact that candidate matches may not be generated, i.e. false negatives. This can be due to the fact that the match is not present in the reference set or is present but is not detected. The former occurs when the reference set based on scientific and technology fields is unduly narrow. Several potential matches may be lost when a patent (publication) by an author (inventor) present in the reference set is not included in the reference set due to technology (science) classification restrictions or errors. For example, an inventor of a patent in the nanotechnology field might have relevant publications in a field that is not included in the definition of nanotechnology publications (Meyer, 2005). However, these “uncharacteristic” publications (or patents) might have important performance-related effects, especially in emerging scientific and technology fields such as nanotechnology. Conversely, candidate matches might not be generated, even though the match potentially exists in the reference set, if due to naming conventions or mistakes in the corresponding databases the author-inventor name match is not detected. Rather than searching for any possible mistake or naming convention variation, efficiently screening for such missing matches is potentially valuable.

Text-Based Profiling Method

A key improvement of our match-making process is the use of text-mining techniques that go beyond the “simple” coupling of authors and inventors based on their first and last names. We use information contained in publication and patent abstracts and generate keywords to associate with the author of a publication and the inventor of a patent. The novelty of this technique is that keywords are generated through text-mining rather than the more standard use of search strings, which requires the researcher to pre-specify what kind of information he is looking for. In the proposed approach, keywords consisting of one or more terms are typically extracted through weighted frequency counts and subsequent statistical processing. Figure 1 depicts the approach: we focus on the scientific and technological output of a single research institution (A) in a given period and extract keywords that are *common* to both patents and publications. After a series of text processing operations such as “stop” word elimination, noun phrase detection and stemming, a domain vocabulary, or thesaurus, is obtained (B). This thesaurus is used to automatically index both datasets (C). By means of these indices and a simple vector-based similarity computation, each patent is subsequently compared with all publications in the database and the best-scoring (publication) hits are considered relevant (D). The final step connects the inventors named in each patent with the authors of the most similar publications (E). We shall now look at each of these steps in detail.

Figure 1

Overview



Data selection

We apply and validate this method in the specific case of the Interuniversity MicroElectronics Centre (IMEC) in Leuven, Belgium. IMEC is a leading research and development (R&D) centre in the areas of microelectronics, nanotechnology, design methods and technologies for ICT systems. Its mission is to carry out R&D programs which are 3 to 10 years ahead of today’s industrial needs through European and worldwide collaboration with industry and academia. IMEC was founded in 1984 by the government of Flanders as an independent R&D institute. The case of IMEC should bias our results in favour of finding inventor-author links, as the semiconductor sector is known for its lively interaction between science and industry. As indicated by Meyer-Krahmer and Schmoch (1998), information technology, in particular data processing and

semiconductors, is a science-based technology and a favorable environment in which to find such inventor-author links. IMEC, as a research centre, aims precisely to bridge the gap between science and industry. In this study we cover 261 EPO patent applications (1985-2002) and 1315 scientific publications (1991-2002) appearing in ISI-WoS (see A in Figure 1). We eliminated entries without abstract and retained 260 patents and 1262 publications.¹

Domain vocabulary or thesaurus

A first step in processing each text document (both patent and publication abstracts) is to extract terms, as they form our basic units of analysis. Morphological variants (such as plurals, inflections, etc.) are canonized using the Porter stemmer (Porter, 1980). Noun phrases, i.e. nouns consisting of two or more terms, are detected by withholding recurrent term patterns. We set the threshold at 3 and 4 for patent and publications, respectively. No explicit efforts are made to deal with acronyms and abbreviations or to recognize proper names. Also, although potentially important in mapping different spaces, ambiguities such as synonyms are not considered in this pilot as we wish to assess the performance of an elementary approach.

Applying this procedure to patents and publications results in two separate domain vocabularies. The intersection between these two thesauri forms the controlled vocabulary, which contains the terms that will be used to relate technology and science in this exercise (B in Figure 1). This final vocabulary is used to construct the text indices for both patents and publications (C in Figure 1).

Automatic text indexing

We adopt the common vector space model to encode a document in a k -dimensional term space, where each component w_{ij} represents the weight of term t_j in document d_i . The grammatical structure of the text is hereby neglected and so this is also referred to as a “bag-of-words” representation. The set of all terms t_j is our controlled vocabulary or thesaurus. Each document is indexed for the occurrence of each term in the thesaurus (C in Figure 1). In the document’s index vector we store non-zero weights of the appearing terms, according to the TF-IDF term weighing scheme:

$$w_{ij} = f_{ij} \log\left(\frac{N}{n_j}\right)$$

where f_{ij} is the number of occurrences of t_j in d_i and is referred to as term frequency (TF). N represents the total number of documents and n_j is the number of documents containing term j in the collection. The logarithm is called inverse document frequency (IDF).

Similarity-based linking of patents and publications

The similarity between pairs of documents d_{i1} and d_{i2} is expressed as the cosine of the angle between the corresponding vector representations, as introduced by *Salton*:

¹ EPO patents where “IMEC” or the “Interuniversity MicroElectronics Centre” appeared as the (co-) assignee were retained from the EPO database. Similarly, publications where “IMEC” or the “Interuniversity MicroElectronics Centre” featured as one of the institutions of the authors were withheld from the ISI-WoS publication database.

$$\text{sim}(d_{i_1}, d_{i_2}) = \frac{d_{i_1} \cdot d_{i_2}}{\|d_{i_1}\| \cdot \|d_{i_2}\|}$$

The underlying hypothesis states that high similarity equals strong relevance. Apart from this type similarity matching, well-constructed text indices empower the application of a wide variety of statistical and data-mining techniques, including clustering and classification. We refer to Glenisson et al. (2005) for an illustration and to Bassecouard and Zitt (2004) for an overview.

So for each (*patent, publication*)-pair a score is computed that reflects the *cosine*-based similarities between patent and publication-indices respectively (D in Figure 1). We hereby perform 328,120 (260 by 1262) comparisons and obtain, for each patent, a ranked list of all 1262 publications. Notice that, in this approach, no name information is being used. As such this can be seen as a rather “brute force” or blind approach to linking inventors and authors. Indeed, at least some commonalities between the last names and/or first names of authors and inventors can be expected, which could be exploited upfront in the process. However, as our primary aim is to prove the *validity* of a content-based matching procedure, the blind approach adopted here is a useful baseline for our purposes.

Linkage of inventors and authors

After computing the vector-based similarities for each patent, the final step involves connecting inventors named on the patents to the authors of the most relevant publications. The number of top-ranking publications we have to retain for each patent is not clear a priori. We decide to scan the method’s performance on the entire range of thresholds $n = [1..1262]$ for all patents (E in Figure 1). Note that the limit case, $n = [1..1262]$, corresponds to linking every patent with every publication, which would produce 5,357,434 linkages.

Up to this point, the inventor-author linkages are based merely on content and not on similarities between the people’s names. Therefore, we include a pruning step. A simple heuristic requiring the inventor and author to share their last name is implemented.²

This completes our procedure, which yields a *list of candidate persons* who share not only their last name, but have also produced documents in both technological and scientific areas and this with a relationship threshold n .

The method can be summarized as follows:

1. Associate with each patent the n most relevant publications
2. Connect all inventors on each patent to all authors on associated publications in 1.
3. Filter these inventor-author linkages by :

$$\text{Last_name}(\text{inventor}) = \text{Last_name}(\text{author})$$

² As we are using a small sample with cleaned names in this step we do not face the typically tedious problem of correct name matching between inventors and authors. Implementing our text-based profile method in conjunction with standard name searches will for some cases require a more sophisticated name matching heuristic between inventors and authors for improved precision of such a hybrid protocol. We touch upon this issue in the discussion section.

In the validation phase we check the candidate list we have generated with our reference database (see below).

Validation

Our *author list* is generated from all authors that feature on publications where at least one author has an IMEC affiliation. Our *inventor list* includes all inventors that feature on patent applications that have been (co-)assigned to IMEC. Note that authors and inventors generated following this procedure are not necessarily members of IMEC. The validated data³ at our disposal consists of a reference database with 206 manually checked author-inventor instances of IMEC personnel. An example is provided in Table 1. These 206 instances correspond to 142 unique full author names.

Table 1

Example of manually matched inventors and authors in our reference database. In this reference database only inventor-author entries are matched. No normalization was done.

Full author name	Inventor last name	Inventor first name
Baklanov, M	Baklanov	Michael R.
Baklanov, MR	Baklanov	Michael R.
Baklanov, M	Baklanov	Mikhail
Baklanov, MR	Baklanov	Mikhail
Baklanov, M	Baklanov	Mikhail Rodionovich
Baklanov, MR	Baklanov	Mikhail Rodionovich

For different settings of n , candidate lists are produced following the described procedure and subsequently compared to this reference database. More specifically, we count the number of *unique full author names* that are present in the reference database of inventor-author links, on the one hand, and in the generated candidate list, on the other. As a quantitative evaluation measure we use the percentage of unique full author names in the reference database that are found by our approach (with parameter n). Unfortunately, we do not have a mapping from unique full author names to unique (physical) persons. Currently we are collecting this data with the help of IMEC internal records. We also note that unique full inventor names could have been used as a reference point.

³ Provided by a third party (not IMEC). Name searches on patent inventors and publication authors were performed and validated with the IMEC personnel list. Unfortunately, the validated reference set does not include all instances of inventor-author links at IMEC, as our method revealed by uncovering additional interesting candidate inventor-author matches not present in the validated data. Therefore, while correct, the validated reference set is incomplete. However, with a proof of validity of the method in mind and given its satisfactory quality, we decided to keep it as our reference set.

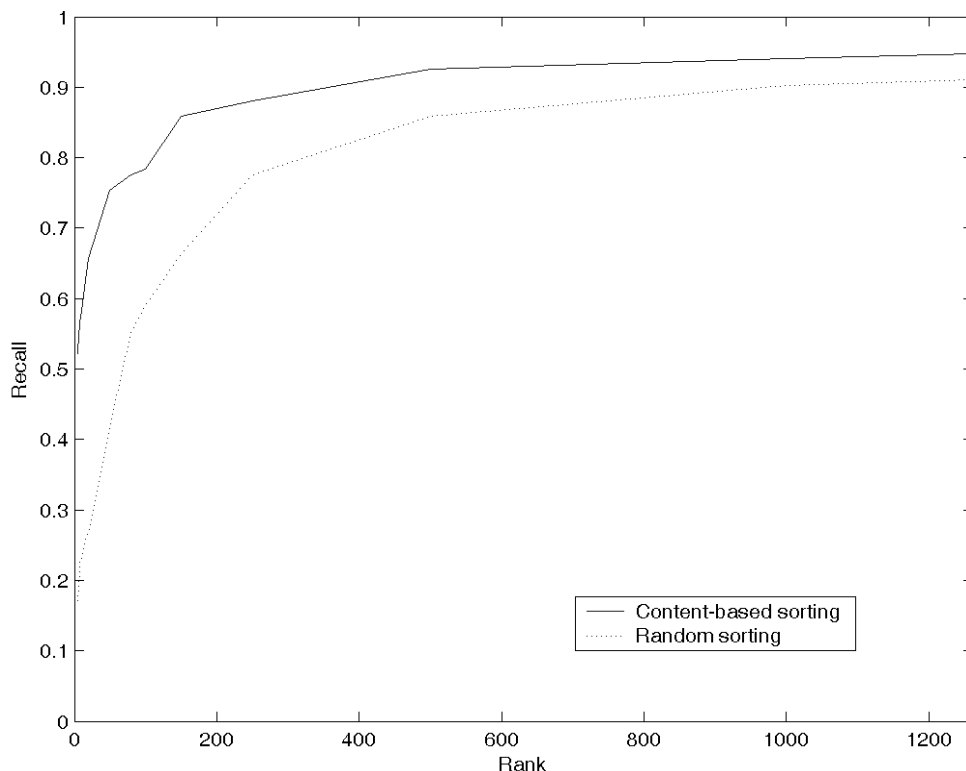
RESULTS

Content-based matching outperforms random matching

In Figure 2 we show the performance of our approach when sorting the publications randomly with respect to the patents. The x -axis depicts the threshold n , i.e. the number of publications included in decreasing order of similarity. The y -axis represents a *recall* measure, i.e., the percentage of overlap in unique full author names between the candidate list and the validated reference database. From this Figure we can see that including the top 20 matches according to abstract content retrieves 66% of the validated author-inventor links. A similar exercise using top 20 random matches yields a recall of 27%. The recall should be complete (100%) in the limit case where all publications are assigned to each patent (Rank=1262). This is not the case, as some validated links were based on fields with an empty abstract. Similarly, one expects that recall of the random sorting should coincide with observed recall of content-based sorting in the limit case (Rank=1262). The observed discrepancy is due to the adopted randomization procedure, which allows publications to be assigned more than once to one patent (i.e., randomization with replacement).

Figure 2

Recall of validated author-inventor links by increasingly linking publications to patents. IMEC publications are ranked with respect to IMEC patents through cosine-based lexical matching (full line) and random ranking (dotted line).



With the setting $n = 20$ we also obtain 261 “hypothetical” links (i.e. candidate links not present in the reference database), corresponding to 59 unique author names. An example is given in Table 2. The example fulfills all the conditions of being an IMEC employee that is active both

as an author and as an inventor. This illustrates that our adopted validation set is incomplete, but above all it strengthens the argument in favor of our approach, as we are able to detect instances that are overlooked by a manual approach.

Table 2

Example of a candidate match not present in the validation set.

Hypothetical Matches				
Name_author	First Name	Last Name	EP	UT
De Ceuninck, WA	Ward	De Ceuninck	EP 0907085	000072771600009
De Ceuninck, WA	Ward	De Ceuninck	EP 0733909	000072771600009
De Ceuninck, WA	Ward Aimé Stefan	De Ceuninck	EP 0395149	000072771600009

Discussion

Given our goal of establishing a proof-of-concept in this first stage, the Recall plot measures the stand-alone power of text-based profiling in matching an inventor with possible authors. Linking an individual patent with the 20 “closest” publications results in a 66% success rate of matching the inventor with the same person as an author from a validated (but incomplete) database of IMEC inventor-author links. As our reference database has been cleaned for author and inventor names, a simple string search on last names would have provided us with a 100% hit rate. Therefore, the performance of text-based profiling should not be measured in absolute terms. The ultimate goal of this method is to use it in conjunction with name searches by eliminating false positive matches or by generating additional candidate matches that existing matching methodologies based on name searches might have missed. The relevant performance measure, therefore, is the success rate *relative* to a random matching of publications with the patent in question. Indeed, we want to determine the power of a text profile to rank publications based on their lexical similarity with respect to a set of given patents. Here, our method performs considerably better, with 66% recall, versus only 27% for random matching of publications with a specific patent.

Considering that our ultimate objective is to apply this method to large-scale databases, we believe that the relative performance will *improve* when more and less-related publications are added to our universe of publications to consider. We argue that IMEC as an institution already would represent a coherent set of publications. As a result, the random assignments of publications to patents are, in fact, not that random, considering the restricted universe they are drawn from in this pilot study. The addition of new, less related publications is therefore likely to improve the performance of our method.

Our method of text-based profiling can help to discard false positives in the extreme but not uncommon case of homonyms, where authors share their last name but have a different text profile. Validating inventor-author links in studies using the personnel list of a university as a starting point has been a challenge because of homonyms. The information needed to discriminate between true links and false positives consists basically of country/address information. Since homonyms are likely to surface in the same country and regions, it is unlikely that location information provides a good way to discriminate between different people as inventor and author. Text-based profiling information might discriminate better. We plan to add publications of randomly selected authors with the same last name as inventors featured in our

reference database in order to quantify the performance of text-based profiling in terms of precision and recall. Nevertheless, we note that two different people with the same last name publishing on similar topics will be hard to distinguish even with text-based profiling.

More important than rejecting false positives, our method gives us the opportunity to generate candidate matches that current methodologies would fail to detect, i.e. false negatives. First, many studies artificially limit the initial set of patents and publications for selecting inventors and authors, subject to matching to predefined science and technology classes. However, there is little guidance in matching patents and publication classes that are likely to cover all the potential for inventor-author linkages. Text-based profiles allow more flexibility in the construction of a reference set of patents and publications. At the extreme, author-inventor name matches generated by pure brute force name searches over the full patent or publication databases, irrespective of science and technology classification, can be qualified by text-based profiles of the matched inventor and author.

Second, our method can generate candidate inventor-author matches where names are similar but not identical. Given the frequent coding mistakes in individual databases and different coding procedures across databases, a potentially significant number of candidate matches between inventor and authors names are lost. Our method allows for a more efficient search of approximate matches by combining name similarity and text profile similarity, rather than scanning the complete databases with name variations (e.g. Van Raan vs Raan, v).

Conclusion

In this pilot study we examine the performance of text-based profiling in recovering a set of validated inventor-author links. In a first step we match each patent in our data with a predetermined number of publications based on their similarity in content. Next, we compare the inventor names on the patent with the author names on the highest ranked matches for the occurrence of name matches. Finally, we compare these candidate matches with the names listed in a validated set of inventor-author names. Our text-based profile method performs significantly better than a random matching of patents and publications, suggesting that text-based profiling is a valuable complementary tool to the name searches used in previous studies. We are working to improve and extend our method in different directions.

First, as Meyer-Krahmer and Schmoch (1998) indicate, collaborative research and informal contacts are very important in microelectronics. In particular, these authors find that conferences are more important than publications in characterising the interactions between universities and industry. Furthermore, Glänzel et al. (2006) show that, for engineering disciplines, publications in conference proceedings amount to over 40% of measured scientific activity. Therefore, we believe that including conference proceedings, now also available in the ISI-Thomson database, might improve the generation of candidate inventor-author links. Extending our method to the proceedings data is straightforward, but the validated reference set does not include proceedings as part of the publications.

Second, in order to evaluate the improved precision of our method with respect to generating *all* existing inventor-author links at an institution such as IMEC from the accessible large-scale patent and publication databases, we are currently collecting a *complete* and validated set of persons involved in patenting and publications from IMEC's internal records. Precision would be defined as the percentage of inventor-author links from the validated complete set we

capture using name searches augmented with text-based profiling, compared to a search based on names only.

In a following step, our method will be used to create text-based profiles of persons rather than merely to link an inventor of a patent with an author of a publication. Patents can be linked with other patents of the same inventor, and publications can be linked with other publications by the same author. Such inventor profiles can be matched with author profiles to construct a more complete profile of different persons acting as brokers on the science-technology exchange.

Finally, one of the ultimate purposes of developing this method is to track researchers who cross the boundaries of science and technology. Meyer-Krahmer and Schmoch (1998) report that knowledge exchange and personnel recruitment rank as the most important objectives for industry in interacting with universities in microelectronics. As knowledge moves with people, examining the mobility of researchers between academia and industry therefore seems fundamental for a better understanding of industry-science links. The case of IMEC provides us with an interesting environment in which to track these movements. As part of its interaction with industry IMEC has organized formal residencies for personnel of affiliate companies. After working and collaborating at IMEC for a predetermined time, developing “background” knowledge which can result in joint publications or joint patents, these residents move back to their sponsoring company to develop “foreground” knowledge, i.e. commercially viable knowledge and technology. The challenge is to uncover links between such developed foreground knowledge and background knowledge from IMEC, based on text-based profiles of inventor-authors who passed through IMEC and show up at a later stage as inventors on private company patents.

While at this point we have only presented preliminary results on the potential of text-based profiling – a proof-of-concept – we are convinced that it is a valuable complementary tool to the extended name searches used in previous studies and are in the process of developing several interesting applications.

References

- Bassecoulard, E. and M. Zitt, 2004, Patents and publications: the lexical connection. In Moed et al., *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems*. Dordrecht: Kluwer Academic Publishers, 665-694.
- Etzkowitz, H. and Leydesdorff, 1997, *Universities in the global knowledge economy: a triple helix of university-industry-government relations*. London: Cassell Academic.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzmann, S., Scott, P. and M. Trow, 1994, *The new production of knowledge: the dynamics of science and research in contemporary societies*. London: Sage.
- Glänzel W., Schlemmer B., Schubert A. and B. Thijs, 2006, Proceedings Literature as Additional Data Source for Bibliometric Analysis. Forthcoming in *Scientometrics*.
- Glenisson P., Glänzel W., Janssens F. and B. De Moor, 2005, Combining full-text and bibliometric information in mapping scientific disciplines. Forthcoming in *Information Processing and Management*.
- Meyer-Krahmer and U. Schmoch, 1998, Science-based technologies: University-Industry interactions in four fields, *Research Policy*, 27, 835-851.
- Meyer M., Siniläinen and J.T. Utecht, 2003, Toward hybrid Triple Helix indicators –A study of university related patents and a survey of inventors, *Scientometrics*, 58(2), 321-350.
- Meyer, M., 2005, Inventor-Authors: Knowledge integrators or weak links? An exploratory comparison of co-active inventing peers in nano-science and technology, working paper 2005/1, Helsinki University of Technology, mimeo.
- Moed, H.F., Glänzel, W. and U. Schmoch (eds.), 2004, *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems*. Dordrecht: Kluwer Academic Publishers.
- Noyons, E.C.M., Van Raan, A.F.J., Grupp, H. and U. Schmoch, 1994, Exploring the science and technology interface – inventor author relations in laser medicine, *Research Policy*, 23(4), 443-457.
- Porter, M.F., 1980, An algorithm for suffix stripping, *Program*, 14(3), 130-137.
- Rabeharisoa, V., 1992, A special mediation between science and technology: when inventors publish scientific articles in fuel cells. In Grupp, H., *Dynamics of science-based innovation*. Berlin: Springer, 45-72.
- Rapmund, A., Gulbrandsen, M. and E.J. Iversen, 2004, Academic patenting in Norway, mimeo.
- Schmoch, U., 2004, The technological output of scientific institutions. In Moed et al., *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems*. Dordrecht: Kluwer Academic Publishers, 717-731.
- Van Looy, B., Callaert, J. and K. Debackere, 2006, Scientific and Entrepreneurial activities within Academia: Mutually reinforcing, hampering or just co-existing? Forthcoming in *Research Policy*.

Verbeek, A., Andries, P., Callaert, J., Debackere, K., Luwel, M. and R. Veugelers, 2002b, Linking Science to Technology, Bibliographic References in Patents – Science and Technology Interplay: Policy relevant findings and interpretations. European Commission, Brussels, EUR 20492, Vol. 1.